

DOCUMENT RESUME

ED 034 710

SP 003 380

AUTHOR Wright, Calvin E.
 TITLE Special Problems of Evaluation Activities in an Individualized Education Program.
 INSTITUTION Westinghouse Learning Corp., New York, N.Y.
 PUB DATE 69
 NOTE 7p.; Paper presented to the American Psychological Association, September 1, 1969, Washington, D.C.
 EDRS PRICE EDRS Price MF-\$0.25 HC-\$0.45
 DESCRIPTORS *Evaluation Needs, *Individualized Instruction, *Measurement Goals, *Standardized Tests, *Testing Problems
 IDENTIFIERS Project PLAN

ABSTRACT

A project designed to determine the relevance of existing standardized achievement tests to the goals of individualized instruction is among the ongoing research activities of Project PLAN, a computer-supported individualized education program. Standardized achievement tests (including the Metropolitan Readiness Test, the Scholastic Aptitude Tests, and the Iowa Tests of Basic Skills and of Educational Development) were administered in the fall of 1968 and the spring of 1969 to control group students and students enrolled in the PLAN program, in grades 1, 2, 5, 6, 9, and 10. Comparative analysis of the data has not been completed, but preliminary results indicate little or no significant difference between the two groups of students in terms of the limited number of instructional objectives these tests are designed to measure. The experiment supports the contention that standardized tests are inadequate for a comprehensive evaluation of a program of individualized instruction. Results also suggest the need for the development of a new series of achievement tests adapted to specified and expanded instructional objectives of both PLAN and control classes in order that the differences between the two can be more thoroughly analyzed and more effectively evaluated. (JS)

ED034710

S

OCT 20 1969

Special Problems of Evaluation Activities in an

Individualized Education Program¹U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

Calvin E. Wright

Westinghouse Learning Corporation²THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

Evaluations and research activities in Project PLAN can be categorized into some seven major areas. These areas include: evaluation of the accomplishments of students; matching students to the learning materials, such as suggesting programs of studies and individual Teaching-Learning Units based upon student characteristics; evaluation of PLAN materials; evaluation of instructional objectives; evaluation of PLAN teachers; evaluation of PLAN testing instruments such as the module tests and other achievement and developed abilities tests; and overall evaluation of the system including the computer support.

In process and pending evaluations of the accomplishments of students include the assessment of progress on individualized programs of studies, the number of TLU's completed, the time taken to complete the TLU's, scores received on the module tests by success category, the number of objectives passed per module, and similar statistics. Also available for analysis are behavior observation records and evaluative judgment data, background and biographical information, academic grades, and scores on commercial and PLAN developed aptitude, ability, and achievement tests. Soon to be added are tests of interest, self-knowledge, motivation, and responsibility. To the extent possible, data are collected not only on PLAN students, but also on groups of students in traditional classrooms that have been designated as controls. For the larger PLAN schools, the controls are either all of the non-PLAN students in

¹Paper presented to the American Psychological Association, September 1, 1969, Washington, D. C., as part of a symposium on Project PLAN: A Computer-Supported Individualized Education Program.

²Now with the American Institutes for Research, Palo Alto, California

S1003380

the same grade level, or a sample thereof. For those schools where all the students at a given grade level are in the PLAN program, a Control group was designated at a comparable school in the same district.

There are a multitude of special problems of evaluations activities in an individualized education program. Today, however, I would like to discuss only some of the problems that relate to evaluations based upon standardized tests.

To begin with, one may question the appropriateness of the use of standard achievement tests for evaluating a program of individualized instruction. Such tests are usually closely attuned to what is being taught in the traditional classroom, where all students are exposed to the same subject matter at the same time, regardless of whether or not each of them has mastered the previous assignments. When tested, they are all at relatively the same point in their studies.

In a PLAN classroom, on the other hand, each student proceeds at his own rate toward the attainment of his educational objectives. When he believes that he has mastered the objectives of the Teaching-Learning Unit upon which he is working, he is free to challenge the corresponding module test to prove his mastery. If he does not pass the test, he must review or restudy the material until an appropriate level of proficiency has been achieved. Only then can he go on to his next assignment. At the time a standardized test is administered, some students in a PLAN class would be considerably ahead, others considerably behind, the students in the traditional classroom. To the extent, however, that the major instructional objectives of the PLAN and conventional classrooms are the same, results on standardized tests should be considered. Although not adequate for a comprehensive evaluation, they do provide some information.

It is well known that a typical research design involves administering some test to both an experimental and a control group at two different points in time, and comparing the results obtained. Hopefully the two groups are relatively comparable on the pre-test, which is usually administered before the experimental group receives

whatever special treatment is to be assessed. In a continuous project in an educational setting, the basic requirements of this research design are difficult to fulfill, especially after the first year. For example, in project PLAN, an attempt was made to select the experimental students randomly from among all students at a given grade level at each school. However, final choices were made by the participating schools from among the students selected, after consultation with the parents involved. There could have been a tendency to choose either the more able students, or to select the less able on the assumption that they would benefit more. In those situations where Control schools were designated, there was no way in which the comparability of the two groups could be assured.

Another set of problems relate to the selection of the testing instruments. Most schools have their own testing programs which differ from district to district, and there is a reasonable reluctance to administer additional tests. Once a series of tests has been selected, it will probably still be necessary to match tests from one battery to another before a comparison can be made. For example, the Stanford Achievement Tests Primary I battery appropriate for the beginning of grade 2 has somewhat different tests than the S.A.T. Primary II battery that is used at the end of grade 2.

The scores received from most standardized achievement test batteries are usually expressed as grade equivalences, although for certain purposes raw scores might be more useful. With grade equivalencies, however, the relative standings of the groups can be determined by a comparison with expected grade placement at the time of testing. An evaluation of growth between test administrations is usually made, with the hope that the publisher's norms were adequate in determining the grade equivalence scores. In any case growth scores should be interpreted cautiously. It has been pointed out by some authors, for example, that growth may not be uniform throughout the school year (e.g., Beggs & Hieronymus, 1968). Other writers have pointed out numerous problems in the use of any growth or change scores (Cronbach, 1969; Harris, 1963).

One of our in process evaluations projects is a comparison of PLAN and Control student results on standardized achievement tests administered in the Fall of 1968 and the Spring of 1969. To the extent possible an attempt was made to choose those test batteries that were routinely used by the local school districts. The Metropolitan Readiness Test was selected for Fall testing in grade 1, the Stanford Achievement Test for the end of grade 1 and for grade 2. The S.A.T. was also chosen for administration in the Western schools in grades 5, 6, 9, and 10. Eastern PLAN school districts were asked to give the Iowa Tests of Basic Skills at grades 5 and 6, and the Iowa Tests of Educational Development at grades 9 and 10.

The analyses of these data have not been completed, but preliminary results are available for grades 1, 2, 5, and 6. For grade 1, test scores were received for about 300 PLAN students and a somewhat larger number of Control students. (The specific number of students differed from test to test within the same battery since some scores were missing for some students.) In the Fall, PLAN and Control students had about the same mean on the Metropolitan Readiness Test total score, but with the PLAN group slightly but non-significantly higher. The means were at about the 90th percentile on national norms. In the Spring, both groups were about equal on S.A.T. Word Meaning, Paragraph Meaning, and Word Study Skills. PLAN was one month ahead of the Controls on Arithmetic, while Controls were about two months ahead on Spelling. The largest difference was for Vocabulary, where the PLAN group was ahead about four months. Data were not available to compute tests of significance for these preliminary analyses, but it is doubtful that the observed differences were significant. Both PLAN and Control students were above expected grade placement, as much as seven months for both groups on Word Study Skills, and nine months on Vocabulary for the PLAN group.

There were some 400 PLAN and 400 Control students tested with the S.A.T. in the Fall and again in the Spring in grade 2. On the tests in common to both Fall and Spring testings, both PLAN and Control were about equal in the Fall, with observed differences no more than one month. From Fall to Spring, growth was slightly greater

for the PLAN students on Word Meaning and Paragraph Meaning, the groups were about equal on Spelling, and the Controls had slightly greater growth on Arithmetic and Word Study Skills. The latter difference, the largest found, was only three months. Growth equalled or exceeded that expected in the six months between the two test administrations except for Word Study Skills and Arithmetic for both groups. All scores also exceeded expected grade placement, as much as nine months for the PLAN students on Word Study Skills, the test for which smallest growth was found.

Grade 5 data were received on the S.A.T. from a total of about 250 PLAN students and 300 Controls. In the Fall, mean scores for PLAN students were two to five months greater on all tests except Arithmetic Computation, where the Controls were about two months ahead. During the year the growth for PLAN students was equal to or greater than that for the Controls for all tests except Arithmetic Computation. The largest difference, however, was only three months for Science. Neither group had the expected six months growth on Word Meaning, Paragraph Meaning, or Social Studies, and the Controls showed a one month loss on this latter. PLAN students also had less than six months growth on Arithmetic Computation, as did the Controls on Spelling, Arithmetic Concepts, and Science. The greatest growth was in Language: eight months for PLAN, six months for Control students. In the Spring, PLAN students were at or above grade placement on all tests except Arithmetic Computation, on which they were nine months below expectation. Control students were seven months below grade placement in Arithmetic Computation, but in addition were below on all other tests except Arithmetic Concepts, where they were about at the norm.

S.A.T. results for Grade 6 were received for some 190 PLAN and 300 Control students. In the Fall, both groups had about the same mean scores on Word Meaning and Paragraph Meaning. The PLAN group was three to four months ahead on Spelling, Language, Arithmetic Concepts, Arithmetic Applications, Social Studies, and Science, while the Controls were about four months ahead on Arithmetic Computation.

Growth for the grade 6 PLAN students ranged from no growth on Arithmetic Applications to one year and one month on Word Meaning. Growth was greater than six months on all tests except Arithmetic Applications, Arithmetic Concepts, Social Studies, and Science. For the Controls growth ranged from no growth on Paragraph Meaning to nine months for Arithmetic Computation. They showed growth greater than six months only for Arithmetic Computation, Word Meaning, and Spelling. The PLAN group had three to six months greater growth than the Controls on Word Meaning, Paragraph Meaning, and Language, while the Controls exceeded PLAN two months on Arithmetic Applications and four months on Arithmetic Concepts. The PLAN students were at or above expected grade placement except on Arithmetic Computation, where they were eight months below in the Spring. The Controls were six months below expected placement in Language, and were also below three months or more on Arithmetic Computation, Arithmetic Applications, and Social Studies.

As mentioned before, the results just presented are preliminary, and will have to be refined in light of the various problems that were discussed. Subsequent analyses should probably be done either by school or individual class, and the competency of the teacher in individualizing instruction should be taken into account. If, however, one assumes that the typical standardized achievement test can only accurately measure mastery of a relatively small number of instructional objectives common to many programs, then significant differences between the PLAN and Control groups might not be expected when these instruments are used. This might be investigated by determining how many of the instructional objectives typically mastered by PLAN students in some nominal grade level actually appear in an end of year standardized test.

Another approach that is being worked on is the development of a new series of achievement tests that will test mastery of the specified instructional objectives of both the PLAN and Control classes. When finished, these instruments can be used to determine what both groups have learned, and what one group may have learned that the

other did not. Since the scores from the tests will directly indicate whether or not the students have mastered the objectives, a more adequate evaluation of PLAN can be made.

As critical as the determination of the scholastic achievements of PLAN students may be, equally if not more important will be future evaluations of the extent to which the PLAN program succeeds in its other goals. These goals include assisting students to develop a sense of responsibility for their educational, personal, and social development, and to make realistic decisions and choices so that they may make full use of their talents in their future adult roles. We would be satisfied were these the only goals in which PLAN succeeds.

References

SEP 18 1969

Beggs, D. L., & Hieronymus, A. N. Uniformity of growth in the basic skills throughout the school year and during the summer. Journal of Educational Measurement, 1968, 5 (2), 91 -97.

Cronbach, L. V., & Furby, Lita How we should measure "change" --or should we? Unpublished manuscript, Stanford University, June, 1969.

Harris, C. W., (Ed.) Problems in Measuring Change. Madison: University of Wisconsin Press, 1963